

# NLP/CL: Carnegie Mellon

David Palzer

# Who?

- ▶ Faculty:
  - ▶ Jaime Carbonell (Analysis of Rare Incident-Event Languages)
    - ▶ Identification of key entities and machine translation (low-resource languages)
  - ▶ Chris Dyer (Bridge the “Linguistic Divide” in Natural Language Processing)
    - ▶ Allows for low-resource languages to borrow lexical material from high-resource languages
  - ▶ Alan Black (Dialport: Enabling Spoken Dialog Research with Real Data)
    - ▶ Conduct challenges for real users allowing for data collection and increased interest
  - ▶ Tom Mitchell (Never Ending Language Learner)
    - ▶ System that learns over time to read the web and extract information

# What?

- ▶ Machine Translation
- ▶ Dialog Systems
- ▶ Machine Learning
- ▶ Speech
- ▶ Information Retrieval
- ▶ Language Modeling
- ▶ Big Data

# Things of Interest

- ▶ [NELL](#) - Never Ending Language Learner
- ▶ [LTI](#) - Language Technologies Institute
- ▶ [AVENUE](#) - statistical syntax-driven machine translation, applied to resource-poor (including endangered) languages and resource-rich languages
- ▶ [CHILDES](#) - child language data exchange system (part of TalkBank)
- ▶ [GRASP](#) - parsing the CHILDES corpus
- ▶ [JAVELIN](#) - open domain, multilingual question answering (English, Japanese, Chinese)

# Things of Interest

- ▶ [Minorthird](#) - an open-source Java package of information extraction and text classification learning tools
- ▶ [Noah's ARK](#) - statistical NLP research group
- ▶ [RADAR](#) - building and evaluating a learning cognitive assistant
- ▶ [Scone](#) - knowledge representation for meaning-based understanding of text
- ▶ [SIDE](#) - a summarization integrated development environment
- ▶ [TagHelper](#) - a semi-automatic tool that facilitates reliable content analysis of corpus data
- ▶ [TalkBank](#) - communication datasets and software